天文卫星下传数据处理系统的规划[1]

赵海升 李承奎 贾淑梅 宋黎明
（中国科学院高能物理研究所 北京 100049）

摘要：天文卫星获取的数据需要经过天文卫星下传数据处理系统的一系列加工生成可以分析的数据产品，这些产品及相应软件要发布给国内外用户。同时数据处理系统还要监测载荷状态，数据质量及天体源爆发情况等。这样，下传数据处理系统的建设直接关系到物理成果的获取，规划该系统就变得非常有意义。本文从数据产品定义，子系统规划，数据流程等方面介绍该系统的规划，并提出以模块化开发方式从整体上协调各个子系统的开发，使各个系统及其软件相互配合，共同促进科学产出。

关键字：天文卫星；预处理；数据分析；监测快视
中图分类号：P171.5

0 引言

天文卫星获取的数据需要经过一系列步骤生成可以分析的数据产品，同时在这个过程中，还需要监测载荷状态、数据质量及天体源爆发情况等，连同数据接口等构成了天文卫星下传数据处理系统（简称下传数据处理系统）。卫星发射前，地面需要具备卫星数据处理能力，并在发射后及时的调整软件参数及算法，开展标定工作。这样，规划下传数据处理系统就变得非常有意义。而且目前，爱因斯坦探针卫星（EP）[1]已经正式立项，增强型X射线时变与偏振天文台（eXTP）也在预研阶段，建设这些卫星的下传数据处理系统也将要提上日程。

下传数据处理系统的主要任务包括三个方面：一，发布科学数据产品，这些数据产品为经过标定及初步筛选的数据（1级产品，包括工程数据和科学数据）和经过完整筛选后提取的科学数据（2级产品，包括能谱、光变、图像及能量响应文件等），同时要发布标定数据库，并实时进行标定数据库的更新；二，发布处理数据产品（1级）的软件，提供软件参数及用户使用手册；三，实时的监测载荷状态和数据质量，并对暴发源、暂现源等识别，特别是要发现新源。工作上，下传数据处理系统要完成软件开发，包括软件的需求设计、开发和测试等步骤，其中还要完成数据处理中关键算法的研究，尤其是要完成仪器地面标定工作。本文主要关注卫星发射前下传数据处理系统的建设，包括数据定义及数据流程，各个子系统的功能及接口，以及系统的运行流程等，并介绍系统建立的主要接口，主要工作内容及软件工程化，目的是使得整个系统能够协调工作，促进科学产生，并提高软件开发效率和运行效率。

1 系统设计

1.1 数据定义及数据流程

卫星下传的数据一般经过简单的解帧、分组和去重复后形成一系列数据文件，我们按数据的下传通道将其定义为遥测数据（0T）和数传数据（0L，0级），它们是下传数据处理系统的输入。0L包括卫星平台采集的数据、载荷采集的工程数据及科学数据。科学数据和工程数据的处理是本系统的关键，它们要经过解包、数据转换、文件拆分和合并等预处理步骤，再进行标定处理（比如电子学增益修正，噪声扣除，事例重建等），直至发布。我们定义1U，1C和1L（1级）分别表征这些步骤中产生的数据产品：1U为没有经过标定处理的数据产品；1C为标定后的数据产品；1L为对外发布的数据产品，它一般是1C或者1U的子集。2L（2级）是在1L基础上产生的高级数据产品，它们经过完整的数据筛选和标定，可以直接用于科学分析，比如，通过对能谱的拟合得到源参数等。监测和快视处理步骤一般比较简单，它的输入可以是0T，也可以是0L处理中的一些数据，输出的结果则

1国家自然科学基金青年科学基金项目资助（11503029）
作者简介：赵海升，男，博士，研究方向：天文数据分析及算法. E-mail：zhaohs@ihep.ac.cn

定义为 2M。

0T 数据主要用于监测及快视，时效性比较强，但信息一般比较少，它对地面判断载荷性能，及时发现新源从而制定观测计划有帮助；0L 时效性比较差，但是数据全面，是科学研究的基础。1L 可以是未标定的数据，也可以是标定的数据，比如 HXMT[2][3]发布的为未标定数据，而 swift[4]发布标定数据。在处理上，发布未标定数据的优点在于预处理和分析两个阶段比较容易分开，而发布标定数据则需要在预处理基础上经过部分数据分析软件模块的处理。本文推荐采用后者，即 1L 为 1C 的子集，这是因为：（1）发布的标定数据可以直接用于监测快视，一方面标定的数据可以识别一部分噪声及本底，这对发现新源有好处，另一方面监测需要评估标定参数，为重新标定提供依据；（2）用户不需要重新标定而可以直接提取能谱、光变等，但是特别说明的是如果用户分析的是以往数据，一般也要重新进行标定处理，这是因为标定数据库更新特别频繁；（3）标定处理后会降低数据量，利于传输和分析。

0L 至 1L 中，数据文件需要根据观测计划或者监测结果进行重新组合和切割，形成一个完整观测，还要在此观测中将不同工作模式（数据采集策略，工作频率等）数据分开。需要注意的是正常观测可能被某些情况打断（比如发现新爆发现象等），观测计划需要及时更新，而监测结果（特别是姿态数据）则容易出现缺失的情况。观测前后的调姿过程数据可以保存在此观测中（这样相邻时间的两次观测可能包含一部分相同的数据），也可以单独作为一次观测。

0L 在传输中可能出现数据重复和文件重复，后续处理中需要有去重复的策略。去重复可以放到数据切割和重组之时。对每一次数据下发，按观测计划对数据拆分和重组产生新的观测数据（文件集合一般定义为一次曝光），此时需要记录每个新文件的完整性及有效性，并给出版本编号。如果下次下发数据与本次存在数据重复或者文件重复，首先判断两次文件的完整性及有效性，如果后者好于前者，则覆盖已产生的数据文件。完整性表示该起止时间是否已经齐备（即数据时间是否覆盖曝光时间），而有效性指文件的数据缺失情况（数据丢包及 CRC 校验[5]不通过及它们的影响程度）及正确性（与下发机构的接口，表征文件因某些问题而重新产生等情况，比如软件 BUG 修正）。

1.2 子系统规划

我们参照于 2017 年 6 月发射成功的 HXMT 卫星的下传数据处理系统，并将其改进为四个子系统。图 1 为四个子系统在整个系统中的规划，它们的关系可以通过整个系统的运行体现（具体由其软件实现）。数据接收及发布子系统在整个架构中起到调度作用，一旦数据下传（无论是数传还是遥测），它将调度其它子系统的软件完成数据处理，并将结果进行存储；数据预处理子系统是数据处理的基础，它主要完成对载荷原包数据的提取（解包操作），这一步将是数据分析子系统和数据监测与快视子系统的基础，并产生完整的观测数据；数据分析子系统在数据预处理的基础上完成标定等，将产生的 1L，2L 数据发布给数据接收及发布子系统，供其存储，发布；监测与快视子系统对遥测数据，解包后的载荷状态数据及定标后的数据进行处理，也将结果发布给数据接收与发布子系统。整个运行为文件驱动型。

图 1 下传数据处理系统规划示意图.
Fig. 1 Conceptual view of Data Processing System.

各个子系统将完成软件开发及开发的辅助工作，其主要规划如下：

（1）数据接收及发布子系统

负责开发数据接收及发布软件，主要完成数据接收（0T 和 0L），调用其它子系统软件完成数据处理，监测快视等，并调用数据预处理部分模块将产生的数据进行拆分和重组去重复后进行结果归档，并提供 1L、2L 和 2M 的发布接口。主要的接口包括输入数据的传输策略，数据发布接口等。

该子系统在整个数据下传处理系统中起到调度作用。

（2）数据预处理子系统[6]

负责开发数据预处理软件，主要完成工程数据和科学数据（0L）的数据解析，形成科

学事例及记录，还原完整科学事例并修正部分信息，产生完整的观测数据文件（1U）。主要的接口为软硬件接口，即需要明确硬件数据采集策略和数据结构定义等。

其主要工作还包括数据分级定义及内容格式设计，数据处理及修正算法研究等。

数据预处理还需要完成空间环境变量的计算，这些变量连同卫星的电压、温度等构成物理分析的筛选条件。

（3）数据分析子系统[7]

负责开发数据分析软件，主要从 1U 级产品出发，完成数据标定、筛选及高级数据产品的提取（1C/2L），并提供标定数据库（CALDB）。主要接口为软硬件接口及与科学用户接口，前接口主要侧重数据分析过程，而后者侧重物理选择条件的选取。

主要工作还包括地面数据的分析及标定工作。地面分析需要注意事例等级、噪声扣除、坏像素识别等方面，标定则关注增益修正，能量与能道关系（E-C），能量响应及探测效率等，还要完成 CALDB 的设计及读写等。

数据分析软件需要发布给科学用户，参照 HXMT 情况（参考文献[7]），其按模块开发，这些模块包括数据标定，好时间计算，数据筛选，能谱光变提取，能量响应产生及本底产生模块等，软件还要提供一个 pipeline 模块，该模块主要完成上述模块的所有功能，其产生的数据即为 2L。CALDB 作为分析软件的一部分也需要一并发布给用户。

（4）数据监测与快视子系统

负责开发数据监测、快视及显示模块。主要完成载荷状态监测，数据质量监测，科学快视及结果显示等。

主要工作在于研究数据特征，搜寻新源及提取源位置等。其建设的重中之重在于快速识别物理现象（快视工作），利用智能算法代替人员值班等。

1.3 系统流程规划

系统中起到调度作用的是数据接收及发布软件，它针对每次数据下发调度各个软件完成数据处理。Error: Reference source not found 为该软件的调度情况。当一次数据下发完成时，首先调用数据预处理软件完成 OL 至 1U 过程，此过程中有许多文件需要处理，一般也将此过程分为多个步骤（参考文献[6]）。这些文件可能存在依赖关系，比如对事例文件时间进行解析时，其 UTC 与晶振对应关系[8]存在于工程文件中。处理策略为：记录每次下发信息，包括文件类型、起止时间等信息，对没有依赖关系的文件直接处理，对存在依赖而文件缺少的情况使已有文件挂起等待下次数传（或者利用上次数传结果）。数据文件完整性对该过程影响也很大，一般来讲 1U 是完整的（特别是针对分析中需要的数据文件），这是因为分析软件在对事例文件进行标定和初步筛选中，会使用很多辅助文件，比如温度文件（增益随温度变化）。但是 0L 一次下传的数据可能满足不了完整性要求，比较常见的是存在数据文件缺失，各个文件时间不统一等，处理策略为在文件合并和拆分时，如果这些文件不能覆盖一次曝光时间，则该曝光不能处理，直到达到文件数目和时间方面的要求。如果下次数传文件时间与本曝光时间有重叠，且数据文件有版本更新，则重新产生该曝光文件。1U 经过部分分析软件模块，主要是标定模块，事例筛选模块（条件比物理分析较为宽松）等，处理为 1C。1C 部分生成 1L，1L 经过分析软件的 pipeline 后生成 2L。

0T 的下传次数一般比较多，它在监测及快视中是非常重要的，对每次遥测下传，调度都要产生 2M 结果。数传数据的 0L 至 1U 过程及 1C 的部分数据都要被用于监测及快视，0L 至 1U 过程数据主要用于发现新源，而 1C 主要用于监测标定信息（标定是否需要更新等）。这样监测和快视主要有三个监测点：遥测数据，0L 至 1U 中的数据，1C 数据。对一次数传或者遥测，数据接收及发布软件都需要调用监测和快视软件，三个监测点不会全部使用而是以文件完备为依据使用，特别是数传时，如果未能产生 1C，而只能用 0L 至 1U 的数据。三个监测点的侧重点也有所不同：遥测侧重要监测载荷运行状态及发现新源，0L 至 1U 侧重发现新源，而 1C 侧重标定监测。

1L，2L 及 2M 均是可以发布的。调度将对发布的文件进行读写权限的控制。表格 1 为各个子系统的数据输入及输出。

图 2 数据接收及发布软件调度情况.

Fig. 2The process scheduling of Data Receiver and Distributor Software

数据接收及发布需要建立数据库（比如 mysql 数据库[9]）存储一些比较重要的信息：（1）每次数传及遥测下传数据情况，特别是数据处理中各中间环节产生文件的情况；（2）处理中的一些关键参数，比如时间对应关系，观测信息，探测单元分类信息（坏探测器等）；（3）标定相关信息，特别是反映定标移动的参数。这些信息最好通过监测显示呈现。

表格 1 各个子系统的数据输入及输出，以及需要发布的软件.

Table 1 Input, output and developing software of each subsystem.

子系统名称	数据输入	发布软件	数据输出
数据接收及发布	0L/0T	数据接收及发布软件	ALL
数据预处理	0L	数据预处理软件	1U
数据分析	1U/1L	数据分析软件	1C/2L
数据监测及快视		数据监测及快视软件	2M

2 系统工程化

整个系统的建立取决于数据接口，数据流程，标定和本底方法的确立。

数据下传系统需要确定各种接口及实施方法。接口中最为重要的是软硬件接口，这个接口包括数据采集流程和数据结构定义，这两方面主要与卫星载荷系统和平台系统有关。平台采集的数据，比如轨道、姿态及部分载荷运行参数，多由专门机构设计（比如航天五院），形式比较简单，发布格式比较固定。但是与载荷系统的接口就比较复杂了，数据采集流程决定数据文件的组织类型及数据格式，数据结构定义则直接决定数据产品解包算法等。本系统一般与载荷系统的建设同步进行，所以这部分接口也是变动比较频繁的。过去接口一般通过文档的形式签署和更改，这存在很大的滞后性，如果软硬件接口通过某种系统即写即读（比如数据库+显示，或者类似 ARP 等）就比较容易实施，本文建议采用配置文件的形式对载荷数据进行定义（参加文献[6]），而配置文件作为软硬件接口可以通过某种系统呈现。

数据处理流程，数据标定，本底，空间环境以及仪器性能的研究是本系统重要的工作，也是完成数据预处理，分析软件及监测快视的基本保障。它们的研究一般基于两种数据，模拟数据和地面测试数据。模拟探测器数据采集对研究标定及数据处理有帮助；空间环境模拟对研究本底有帮助；地面也可以依据数据定义单纯的产生载荷事例数据等，可以用于数据处理流程等的开发和测试。地面测试数据一般由载荷系统负责，但是数据分析也需要本系统的参与，这些数据比较接近卫星下传数据，数据的研究对研究数据处理算法，标定仪器性能有极大的帮助，尤其是确定事例噪声，事例等级重建等有关键性作用。

卫星下传系统的规划存在很多不确定性，特别是数据影响因素的不确定性和标定不确定性。这主要来源于强的空间辐射环境（本底）难以估计，载荷的形状及位置可能存在形变，探测特性也容易发生变化等。这决定了本系统需要具备较强的可维护性和扩展性。本文建议采用模块化，数据与代码分离及面向对象的开发模式，并从整体上进行软件开发。

天文卫星的数据处理一般采用模块化方式进行，每个模块都有自己的输入参数，可以独立运行，而模块之间传递的是数据文件。这种设计降低了步骤上的耦合，使得系统有比较强的可维护性，特别是针对一些处理上的非线性操作（主要指相邻事例之间有关联），将这些处理单独作为一个模块可使得对其维护而不影响其它部分。另一方面模块化有利于数据接收和发布软件的调用。数据处理中有许多参数：数据选择条件参数，探测器配置参数，标定参数及本底参数等，其中标定、本底参数更新频繁。软件开发中尽量将代码和这些参数分离，使得两者的维护不相互影响。对于监测快视，数据与显示界面也需要分开。

面向对象的开发一般都有比较好的封装和继承特性，对系统性开发有很好的帮助。图 3 是目前 HXMT 用户分析软件的软件框架图，其顶层（BaseLIB 库）为基本功能集合，包括数据读写类，信息输出类，能谱和光变类，CALDB 读取类，基本数据结构类，算法类等。设计中 BaseLIB 应该重点开发，它可以为所有软件或者载荷服务，我们推荐使用 HXMT 用户分析软件作为参考库开发该库。中间层为面向不同的载荷（或者面向不同软件，LIB（A），LIB（B）等），包括文件系统的组成类（完成 pipeline 等），各种与载荷（或

者软件）有关的数据结构类等，如果软件或者载荷比较简单，可以将此层忽略。再下层的libs 及 modules 分别存放与载荷相关的算法及特殊功能类等，tasks 则为各个模块的实现，它主要完成模块的流程操作。这样设计具有层次性，接口、算法、功能实现彼此分开，而流程函数将它们相互连接，如果某一部分需要改变，不会影响其它部分。这种设计还有一个优点，可以降低软件开发和测试成本，如果软件之间彼此孤立，无论是开发成本还是测试成本都是高昂的，而本设计中，模块测试可以反复使用 BaseLIB 等库，从而增加其可靠性。

图 3HXMT 用户分析软件框图.

Fig. 3 Layered software architecture of HXMT Data Analysis Software.

数据流程部分的软件开发为本系统的重中之重，它包括数据预处理，数据分析。两者的开发（或者设计）顺序也应该为预处理在先，分析再后，pipeline 模块则可以更靠后，甚至在卫星上天之后再提供。数据预处理，分析和监测快视都需要深刻的理解载荷及数据，而数据接受与发布则可以独立于其它三个的开发。

3 总结及展望

卫星下传数据处理系统是卫星地面系统的一个重要组成部分，它负责将卫星下传数据进行逐层处理，并将标定处理后的数据 1L 进行发布，并发布分析软件及 CALDB。同时它还起到监测探测器状态和数据质量的作用，并承担爆发源、新源寻找等重要工作。这些对卫星科学产出起到支撑和促进作用，尤其是在发现新源方面，将直接影响科学产出。本文将卫星下传数据处理系统分为四个子系统，分别为数据接收及发布，数据预处理，数据分析和监测快视子系统，它们协调工作，完成数据接收和发布软件，数据预处理软件，分析软件和监测快视软件开发和部署，并提供 CALDB。卫星下传数据处理系统的开发需要遵循利用逻辑性降低代码开发量，利用脑力代替体力的原则从而完成系统设计和开发。工作上还要重点关注数据处理算法研究和标定工作。

本文没有关注硬件建设，事实上数据存储及计算环境是非常重要的，这对软件的部署和运行有重要影响。

[1]袁为民, 张臣, 陈勇,等. 爱因斯坦探针：探索变幻多姿的 X 射线宇宙[J]. 中国科学: 物理学, 力学, 天文学, 2018, 48: 039502.

Yuan W M, Zhang C, Chen Y, et al. Einstein Probe: Exploring the ever-changing X-ray Universe (in Chinese). Sci Sin-PhysMechAstron, 2018, 48: 039502.

[2]Li Ti-Pei, Wu Mei. The hard X-ray modulation telescope mission [J]. Physics, 2008, 37(09):648-651.

[3] Li Ti-Pei. HXMT:A Chinese High-Energy Astrophysics Mission [J], Nuclear Physics B, 2007, 166:131-139.

[4]Burrows, D. N. et al.: 2000, The Swift X-ray Telescope[R], Proc. SPIE Vol. No. 4140,pp. 64 - 75.

[5] Han Jiawei, Zhang Hongqun.Error correction technology based on CCSDS standard for remote sensing satellite. Chin. J. Space Sci., 2009, 29(1): 112-116.

[6] 赵海升, 葛明玉, 李正恒, 聂建胤, 宋黎明. 一种天文卫星数据预处理方法[J]. 天文研究与技术, 2017, 14(3): 376-381.

Zhao Haisheng, GeMingyu, Li Zhengheng, NieJianyin, Song Liming. A Method of Data Preprocessing for Astronomical Satellite[J]. Astronomical Research and Technology, 2017, 14(3): 376-381.

[7] ZHAO Haisheng, LI Chengkui, LI Xiaobo, NIE Jianyin, GE Mingyu, PAN Yuanyue, SONG Liming. High Energy X-ray Telescope data analysis method[J]. Chin. J. Space Sci., 2016, 36(6): 938-944.

[8] 蔡凡，陈丽. 卫星双模授时设备的关键技术实现[J]. 天文研究与技术，2018, 15(1):46-51.
Cai Fan, Chen Li. Realization of Key Technology of Dual Mode Satellite Timing Equipment.Astronomical Research & Technology, 2018, 15(1): 46-51.

[9] 钟守波，韩波，张彦霞，等. 天文大数据管理工具的设计与实现[J]. 天文研究与技术，2015, 12(4): 510 - 517.
ZhongShoubo, Han Bo, Zhang Yanxia, et al. Design and implementation of s software tool package for managing massive astronomical data[J]. Astronomical Research & Technology, 2015, 12(4): 510 - 517.

Planning of Data Processing System for Astronomical Satellite

Zhao Haisheng,Li Chenkui,JiaShumei, Song Liming
Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

Abstract: The data obtained from astronomical satellite is telemetered to ground and converted into high products by the data processing system. The products as well as these analytic software will be opened to the public. Meanwhile, the data processing system monitors status of payloads and quality of data, gives a quicking look to the data.Thus, planning of data processing system will be helpfulfor improving the scientific output,particularly in detecting new astronomical sources. This work will introduce the system from the definition of data productsat the different stages of processing, conceptual view of this system,design of its subsystems and flow of data, and illustrate some details of processing.We also want to establish a method for software development to reduce the cost of human resources and promote scientific output.

Keywords: Astronomical Satellite;Data Preprocessing;Data Analysis; Monitor and Quicking Look